IEEE Xplore RELEASE 2.0

Home | Login | Logout | Access Information | Alerts | Sitemap | Help

Welcome United States Patent and Trademark Office

AbstractPlus

BROWSE SEARCH IEEE XPLORE GUIDE SUPPORT

◂ View Search Results | Next Article ▸

e-mail printer friendly

Document options

Full Text: PDF (454 KB)

Download this citation

Choose Citation

Download EndNote,ProCite,RefMan

» Learn More

Rights & Permissions

Request Permissions RIGHTSLINK

» Learn More

Model selection for wavelet-based signal estimation

Cherkassky, V. Xuhui Shao
Dept. of Electr. & Comput. Eng., Minnesota Univ., Minneapolis, MN, USA;

This paper appears in: **Neural Networks Proceedings, 1998. IEEE World Congress on Computational Intelligence. The 1998 IEEE International Joint Conference on**
Publication Date: 4-9 May 1998
Volume: 2
On page(s): 843 - 848 vol.2
Number of Pages: 3 vol. xxxvi+2561
Meeting Date: 05/04/1998 - 05/09/1998
Location: Anchorage, AK
INSPEC Accession Number:6040297
DOI: 10.1109/IJCNN.1998.685877
Posted online: 2002-08-06 21:45:07.0

Abstract
There has been a growing interest in **wavelet**-based methods for signal estimation from noisy samples. We compare popular **wavelet** thresholding methods with model selection using VC generalization bounds developed for finite samples. Since **wavelet** methods are linear (in parameters), the VC-dimension of linear models can be accurately estimated. Successful application of VC-theory to **wavelet** denoising also requires specification of a suitable structure on a set of **wavelet** basis functions. We propose such a structure suitable for orthogonal **basis functions**, which includes **wavelets** as a special case. The combination of the proposed structure with VC bounds yields a new powerful method for signal estimation with wavelets. Our comparisons indicate that using VC bounds for model selection gives uniformly better results than other **wavelet** thresholding methods under small sample/high noise setting. On the other hand, with large samples model selection becomes trivial, and most reasonable methods (including **wavelet** thresholding heuristics) perform reasonably well

Index Terms
Inspec
Controlled Indexing
neural nets prediction theory signal processing **wavelet** transforms

Non-controlled Indexing
VC bounds VC-theory model selection noisy samples orthogonal **basis functions** signal estimation **wavelet basis functions** **wavelet**-based methods
Author Keywords
Not Available

References

No references available on IEEE Xplore.

Citing Documents

1 Guest editorial vapnik-chervonenkis (vc) learning theory and its applications,
Neural Networks, IEEE Transactions on
On page(s): 985-987, Volume: 10, Issue: 5, Sep 1999
Abstract | Full Text: PDF (24)

◂ View Search Results | Next Article ▸

Help Contact Us Privacy & Security IEEE.or

© Copyright 2005 IEEE – All Rights Reserve

Model Selection for Wavelet-based Signal Estimation

Vladimir Cherkassky and Xuhui Shao
ECE Dept., University of Minnesota
Minneapolis MN 55455
cherkass,xshao@ee.umn.edu

ABSTRACT

There has been a growing interest in wavelet-based methods for signal estimation from noisy samples. Signal denoising involves calculating discrete wavelet transform (using training samples) and then discarding 'insignificant' wavelet coefficients (presumably corresponding to noise). Various wavelet thresholding heuristics for discarding insignificant wavelets have been recently proposed [Bruce et al, 1996; Donoho, 1993; Donoho and Johnstone, 1994]. These methods are conceptually based on asymptotic results for linear models, but also take into account special properties of wavelet basis functions. Wavelet thresholding represents a special case of model selection; hence we compare popular wavelet thresholding methods with model selection using VC generalization bounds developed for finite samples [Vapnik, 1982]. Since wavelet methods are linear (in parameters), VC bounds can be rigorously applied, i.e. the VC-dimension of linear models can be accurately estimated. Successful application of VC-theory to wavelet denoising also requires specification of a suitable structure on a set of wavelet basis functions. We propose such a structure suitable for orthogonal basis functions, which includes wavelets as a special case. The combination of the proposed structure with VC bounds yields a new powerful method for signal estimation with wavelets. Our comparisons indicate that using VC bounds for model selection gives uniformly better results than other wavelet thresholding methods under small sample/ high noise setting. On the other hand, with large samples model selection becomes trivial, and most reasonable methods (including wavelet thresholding heuristics) perform reasonably well.

1. Estimation of Prediction Risk

Prediction risk is the expected performance of an estimator for new (future) samples. Accurate estimation of prediction risk from the available training data is crucial for the control of model complexity (model selection). Classical methods for model selection (including recently proposed wavelet thresholding methods) are based on asymptotic results for linear models. Non-asymptotic (guaranteed) bounds on the prediction risk based on VC-theory have been proposed in [Vapnik, 1982].

There are two general approaches for estimating prediction risk for regression problems with finite data. One is based on data resampling. The other approach is to use analytic estimates of the prediction risk as a function of the empirical risk (training error) penalized (adjusted) by some measure of model complexity. Once an accurate estimate of the prediction risk is found it can be used for model selection by choosing the model complexity which minimizes the estimated prediction risk. In the statistical literature, various prediction risk estimates have been proposed for model selection (in the *linear* case). All these estimates take the form of:

$$\textit{estimated risk} = g\left(\frac{d}{n}\right)\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \qquad (1)$$

where g is a monotonically increasing function of the ratio of model complexity (degrees of freedom) d and the training sample size n [Hardle, et al. 1988]. The function g is often called a *penalization factor* because it inflates the average residual sum of squares for increasingly complex models. Various forms of g have been proposed in the statistical literature, i.e. Final Prediction Error [Akaike, 1970], Generalized Cross-Validation [Craven and Wahba,1979] etc. All these estimates are motivated by asymptotic arguments for linear models and are applicable for large training sets.

Statistical Learning Theory (SLT) provides an upper bound estimation for prediction risk [Vapnik, 1982; Vapnik, 1995]. For regression problems with squared loss function the SLT bound is:

$$\textit{prediction risk} \le R_{emp}\left(1 - c\sqrt{a_1\frac{h_m\left(\ln\left(\frac{a_2 n}{h_m}\right)+1\right)-\ln\left(\frac{\eta}{4}\right)}{n}}\right)_+^{-1} \qquad (2)$$

0-7803-4859-1/98 $10.00©1998 IEEE

where h_m is the VC-dimension of the set of approximating functions and c is a constant which reflects the 'tails of the loss function distribution', i.e. the probability of observing large values of the loss. The quantities a_1 and a_2 are theoretical constants that need to be empirically tuned for a given class of problems. The above upperbound holds with probability $1-2\eta$ (confidence level of an estimate). For practical use, one needs to set the value of the constants c, a_1, a_2, and the confidence level. As recommended in [Cherkassky and Mulier, 1998; Cherkassky et al, 1996], for regression problems with squared loss function, the good choice is $c=1$, $a_1=1$, $a_2=1$, and $\eta=4/\sqrt{n}$.

Further, in order to use (2) we need to estimate the VC dimension of a set of approximating functions. As discussed in [Cherkassky and Mulier, 1998], accurate estimation of the VC-dimension for nonlinear methods (such as feedforward neural networks) is difficult if not impossible. In this study we use only linear methods, such as fixed wavelet basis functions, where the VC dimension can be estimated by the number of free parameters (degrees of freedom). For example, for linear estimators (with m free parameters), the VC dimension is $h_m = m$.

Making all these substitutions into (2) gives the following penalization factor which we call Vapnik's measure *(vm)*:

$$g(p,n)=\left(1-\sqrt{p-p\ln p+\frac{\ln n}{2n}}\right)_+^{-1} \tag{3}$$

where $p=m/n$. Penalization factor (3) is used for model selection comparisons (reported below) with wavelet estimators.

2. Model Selection for Wavelet Estimators

In signal processing, a popular approach for approximating *known* univariate functions (called signals or waveforms) is to use orthonormal basis functions. A *wavelet* is a special basis function which is localized both in time and in frequency.

Discrete wavelet basis function representation of a signal has the form:

$$f(x,w)=\sum_j\sum_k w_{jk}\psi\left(2^j x-k\right) \tag{4}$$

where the wavelet basis functions $\psi_{jk}(x)=2^{j/2}\psi\left(2^j x-k\right)$ form an orthonormal basis, provided that the mother wavelet satisfies certain requirements (i.e. it has sufficiently localized support and zero mean). As common in signal processing, a signal is sampled at fixed x-locations uniformly spaced on a [0,1] interval:

$$x_i=\frac{i}{2^J} \quad \text{where}$$

$$i=0,1,2,\ldots,2^J-1.$$

Due to orthogonality of wavelets, all wavelet coefficients in (4) can be computed from training samples (x_i,y_i) very efficiently using signal processing techniques by taking the discrete wavelet transform of a signal. This paper only considers the discrete wavelet representation (4) corresponding to *fixed* basis function expansion (i.e., linear estimator). Hence, methods considered in this paper are limited to *low-dimensional* problems, i.e. 1D or 2D signals (common in signal processing).

Recently several authors [Bruce et al, 1996; Donoho and Johnstone, 1994] advocated the use of wavelet methods for signal estimation from noisy samples (called *denoising* in Signal Processing). In signal processing, both the training signal and the future signal are sampled at the same x-locations. Wavelet noise removal works by taking the discrete wavelet transform of a signal (i.e., calculating all wavelet coefficients), and then discarding the terms with small or 'insignificant' coefficients. For example, one can discard wavelet basis functions having coefficients below a certain threshold. Finally, the denoised signal is obtained via inverse wavelet transform. The above procedure for wavelet denoising clearly represents a special case of a standard regression problem. The main distinction is that model selection (i.e., determining insignificant wavelet coefficients) is achieved via heuristic rules which are specifically designed for wavelet basis functions under fixed sampling rate assumption.

There are two major claims regarding potential advantages of wavelet denoising:

(a) wavelet basis functions are more suitable for estimating 'nonstationary' signals.

(b) thresholding methods for wavelet denoising [Donoho and Johnstone, 1994] perform superior model selection.

This study is mainly concerned with the model selection claim (b). However, we observed that with *finite* samples it is the model selection (b) rather than the choice of basis functions (a) that has major effect on accurate signal estimation. In other words, using standard discrete Fourier transform with good model selection would yield better results than using wavelets with mediocre model selection. It appears that most claims regarding wavelet denoising in the wavelet literature [Bruce et al, 1996; Donoho, 1993; Donoho and Johnstone, 1994] are made for large samples, when the task of model selection is easy.

The problem of wavelet noise removal is a special case of model selection, and can be addressed in the framework of Statistical Learning Theory, as explained next. SLT specifies a major inductive principle for estimation with finite data, called *Structural Risk*

Minimization (SRM). According to SRM, one needs to specify some form of complexity ordering on a set of possible models (or approximating functions). Then an element of a structure corresponds to a set of approximating functions of fixed complexity, and all elements can be ordered according to their flexibility to fit the data. Model selection under SRM formulation amounts to choosing an optimal element of a structure which provides smallest generalization bound (2). With wavelets, there are total $n = 2^J$ wavelet basis functions, and we specify a structure on this set as follows. Consider the following structure on a set of all discrete wavelet basis functions $\psi_{jk}(x)$: each element of a structure S_m has exactly m basis functions (wavelets). Note that once m basis functions (wavelets) in S_m are specified, minimization of the empirical risk is trivial (due to orthogonality of wavelets) and amounts to estimation of the wavelet coefficients. The structure on a set of orthogonal basis functions is defined by an *appropriate ordering* of all basis functions. The contribution of an orthogonal basis function to the reduction of risk is proportional to the absolute value of its coefficient in expansion (4). Moreover, with a fixed sampling rate, a basis function's contribution to the reduction of risk also depends on its support. Hence, an appropriate structure on a set of all $n = 2^J$ wavelet basis functions may be defined by ranking all wavelets according to their coefficient value adjusted by scale, $|w_{jk}|\ 2^{-j}$. Each element of this structure S_m consists of the first m wavelets ordered according to their coefficient adjusted by scale, $|w_{jk}|\ 2^{-j}$. Then for model selection, Vapnik's measure (3) for estimating prediction risk is used for each set of wavelet functions S_m.

Next we present empirical comparisons between VC bound (3) applied to the above wavelet structure and wavelet thresholding heuristics. The wavelet thresholding methods, i.e., SURE (with hard thresholding) and VISU (with soft thresholding), HYBRID, MINIMAX and MAD are a part of the WaveLab package developed at Stanford University [Bruce et al, 1996; Donoho and Johnstone, 1994]. This is public domain software available via Internet (the WWW address is http://playfair.stanford.edu/~wavelab). The WaveLab uses symmlet wavelet basis functions (by default), so the comparison was done using symmlet wavelets.

The training data is generated using two target functions, Heavisine and Blocks (shown in Fig. 1) corrupted with Gaussian noise. Note that Blocks signal contains significant high-frequency components, whereas Heavisine signal contains mainly low-frequency components. Training samples x_i, $i = 1,\ldots,128$ are equally spaced in the interval $[0,1]$. The noise is Gaussian with SNR=2.5. For each training sample, all 128 wavelet coefficients were found. Then selection of 'significant' wavelets providing an estimate of a true signal is done using wavelet thresholding methods and Vapnik's measure for model selection using the wavelet structure described above. For each method, its approximation error and model complexity are recorded. Approximation error is measured as L_2 distance between the true signal and its estimate, normalized by the standard deviation of the true signal (NRMS error). Model complexity is measured as the number of wavelet basis functions selected by each method, or degrees-of-freedom (DOF).

The above estimation procedure is performed 300 times using different realizations of random training samples, and the resulting empirical distribution of NRMS and DOF is used to compare the methods. The results are presented using standard box plot notation with marks at 95, 75, 50 and 5 percentile of an empirical distribution. See Figure 2. Visual comparison of estimates provided by each method is given in Fig. 3.

3. Summary and Discussion

The above comparisons and other empirical results (not shown here due to space constraints) suggest that signal estimation based on Statistical Learning Theory outperforms wavelet thresholding methods. This is quite remarkable, since SLT bounds (2), (3) used in this comparison are very general and do not reflect specific signal processing formulation (i.e., fixed sampling rate). In contrast, wavelet thresholding methods are custom-designed for signal processing with wavelets. Our experience with wavelet thresholding methods contradicts optimistic claims in the wavelet literature [Bruce et al, 1996; Donoho, 1993; Donoho and Johnstone, 1994]. The reason is that examples provided in the wavelet literature and wavelet software packages use *large-sample* signals. With large samples, model selection is simple, and most (reasonable) methods give good results. The real challenge is model selection with small (or finite) samples. Recall that 'small sample' problems are defined [Vapnik, 1995] as problems where the model complexity (DOF) is of *the same order* as sample size. On the other hand, for *large samples*, the model complexity (DOF) is much smaller (say, 20 times or more) than the number of samples. In the context of signal denoising, the challenge is to develop a method which automatically selects large number of wavelet coefficients when the true signal complexity is high, and small number of wavelet terms when the true complexity is low. Comparisons presented above illustrate small-sample scenario. With 128 samples, best models use 40-60 degrees of freedom for Blocks signal, and 10-15 degrees of freedom for Heavisine signal (see Fig. 2). Notice that VC-based model selection is capable to determine optimal model complexity for arbitrary signals, i.e. it performs well for both Blocks and

Heavisine signals. Wavelet thresholding methods appear to be tuned to a particular signal type, i.e. SURE does well for Blocks (but fails for Heavisine), whereas VISU does well for Heavisine (but fails for Blocks). With large samples, i.e. 1024 or more for Heavisine signal (at the same noise level), there is no significant difference between most wavelet thresholding methods and SLT-based approach.

Acknowledgement: this work was partially supported by NSF grant IRI-9618167 and by the IBM Partnership Award.

References

Akaike, H. (1970), Statistical predictor information, Annals of the Inst. of Statistical Mathematics, 22, 203-217.

Bishop, C. (1995) Neural Networks for Pattern Recognition, Clarendon Press, Oxford.

Bruce, A., Donoho, D. and H.-Y. Gao (1996), Wavelet analysis, IEEE Spectrum, Oct. 1996, 26-35

Cherkassky, V. and F. Mulier (1998) Learning From Data: Concepts, Theory and Methods, Wiley, N.Y.

Cherkassky, V., Mulier F. and V. Vapnik (1996) Comparison of VC-method with classical methods for model selection, in Proc. World Congress on Neural Networks, San Diego, CA, 957 - 962

Craven, P., and Wahba, G. (1979), Smoothing noisy data with spline functions, Numerische Math., 31, 377-403.

Donoho, D.L. (1993), Nonlinear wavelet methods for recovery of signals, densities, and spectra from indirect and noisy data, Different Perspectives on Wavelets, Proc. of Symposia in Applied Mathematics, I. Daubechies (Ed.), v. 47, Amer. Math. Soc., Providence, RI, 173 - 205

Donoho, D.L. and I.M. Johnstone (1994) Ideal denoising in an orthonormal basis chosen from a library of bases, Technical Report 461, Statistics Dept, Stanford University

Hardle, W., Hall, P., and Marron, J. S. (1988), How far are automatically chosen regression smoothing parameters from their optimum?, JASA 83, 86-95.

Vapnik, V, (1982), Estimation of Dependencies Based on Empirical Data, Springer, NY

Vapnik, V, (1995), The Nature of Statistical Learning Theory, Springer, NY

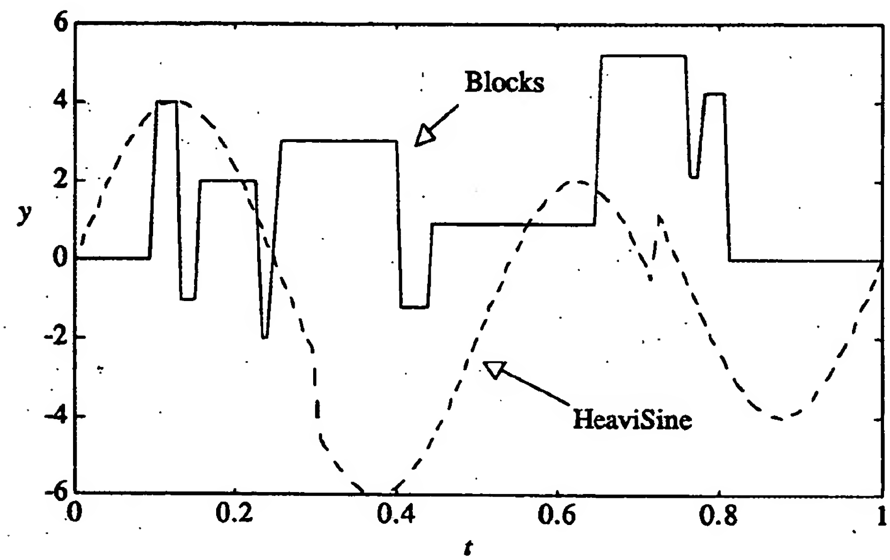

Figure 1. Target functions used for comparisons.

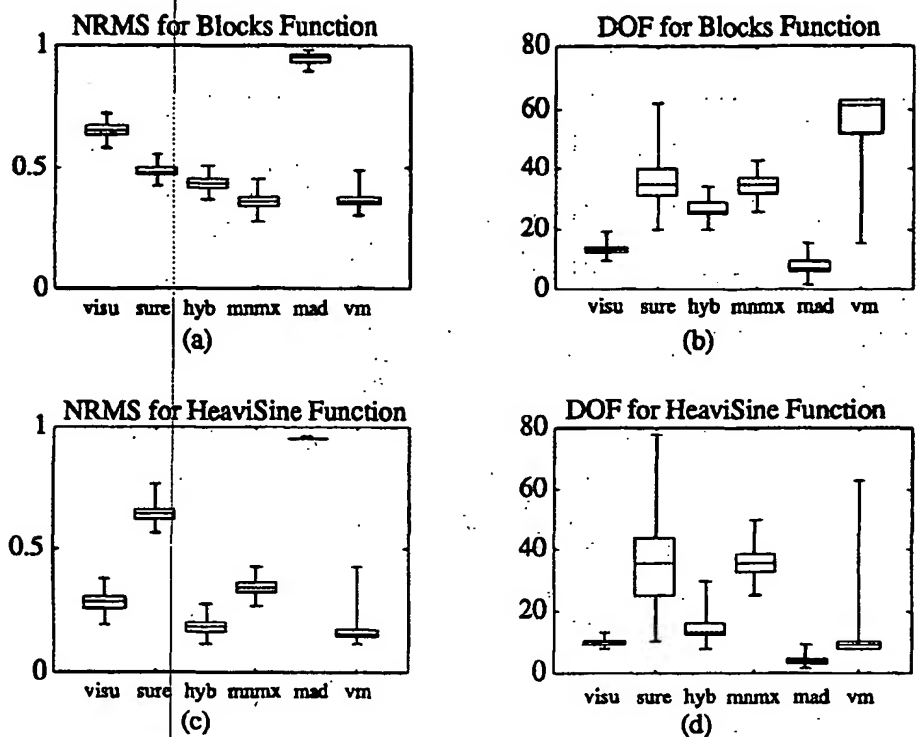

Figure 2. Comparison of methods for wavelet-based signal denoising

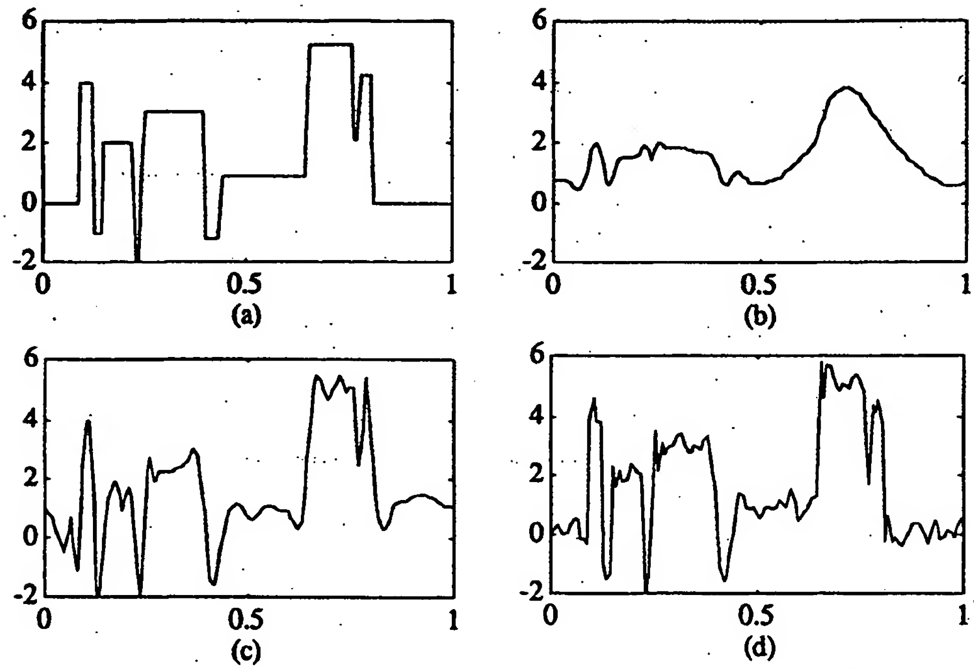

Figure 3. Visual comparison of signal estimates:

(a) The Blocks target function
(b) Fit using Visu procedure
(c) Fit using Sure procedure
(d) Fit using the proposed wavelet structure and VC-bound.